R & D STATUS REPORT

ARPA ORDER NO:	A418
PROGRAM CODE NO:	DO-C9
CONTRACTOR:	David Sarnoff Research Center
CONTRACT NO.:	N00014-93-C-0202
CONTRACT AMOUNT:	$676,870
EFFECTIVE DATE OF CONTRACT:	18 August 1993
EXPIRATION DATE OF CONTRACT:	17 August 1996
PRINCIPAL INVESTIGATOR:	John C. Pearson (609-734-2385)
TECHNICAL CONTRIBUTORS:	John Pearson, Paul Sajda and Clay Spence
SHORT TITLE:	Hybrid Pyramid / Neural Network Vision System
REPORTING PERIOD:	9/1/94 to 11/30/94

1 The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advance Research Projects Agency of the U.S. Government.
2 Ownership in Patent Data included herein is retained by the contractor/subcontractor pursuant to FAR 52.227-12.

Description of Progress:

Pattern Trees and Component-Learning

As part of our goal for automating pattern tree learning, we have tested whether a network can be trained to discover a single salient component which discriminates targets from non-targets (we have previously called this "feature discovery"). Note that this is in contrast to a network which is trained to discriminate target from non-target based on any part of the image of the target. For this test we use the error function

$$\varepsilon_P(\mathbf{w}) = \min_{x \in P}(-\log(y(x, \mathbf{w}))) \qquad \text{(EQ 1)}$$

where P indexes the particular positive example, y is the network output, and $\mathbf{w}$ is the parameter vector of the network. This error function is minimized over a set of positive and negative examples (i.e. targets and non-targets). For negative examples, we divide the non-target regions of the images into parts (squares, except for those which would overlap a positive example, in which case that portion is removed) whose size is the median linear extent of the targets. We consider each such region to be a negative example, and use the average cross-entropy error $(-\log(1 - y(\mathbf{f}, \mathbf{w})))$ for the example's contribution to the total error. We have used this objective function with the building-detection problem, the problem of finding microcalcifications in mammograms (see below), and aircraft detection. In all cases the resulting network detects only part of the target, however, its output is usually very close to one at those examples that it detects, even the false positives, indicating that it is not a good estimate of the probability that a target is present (i.e. the network seems to instantiate a binary decision). This should not be a problem, since we intend to use the output of this network (or a function of it) as an input, representing a meta-feature, to a network which will be embedded in the pattern tree representation.

Learning several components

For the building-detection problem, we trained a second network to find a different salient component than the first. Simply not training the second network on those regions which were detected by the first did not seem to work; the second network was similar to the first and responded wherever the first network responded. A second approach we tried was to use regions classified as targets by the first network as additional *negative* examples for the second network. This approach worked much better, with the second network detecting different locations than the first.

Applications to Biomedical Imagery (Mammograms)

We have applied the neural network/pyramid architecture to the detection of microcalcifications in mammograms (mammogram data provided by Dr. Robert Nishikawa of The University of Chicago). To date, we have trained networks on the

third and second pyramid levels, i.e., at one-sixteenth and one-eighth of the original resolution, using the same oriented energy features as used for the building-detection problem. We compared a non-hierarchical network architecture with our hierarchical detector constructed with two networks, and found that the hierarchical detector performs significantly better. The inputs to the non-hierarchical detector's network were the oriented energies from the zero-th through the third pyramid levels, which are the highest four octaves in the spectrum below the Nyquist frequency. The inputs to the hierarchical detector's second-level network were oriented energies from the zero-th through the second levels (the highest three octaves), plus the outputs of the four hidden units of the level-three network. Thus, the two detectors had the same number of inputs, at the second level. The superior performance of the hierarchical detector is in contrast to the building detector, in which the hierarchical and non-hierarchical detectors had essentially equal performance. One possible explanation is that the hidden units in the building detector network were simply passing information through to higher resolution, without performing any processing needed by the high-resolution network. The hidden units of the third-level net in the microcalcification detector, however, processed information in a way that was useful to the higher-resolution network.[1]

Training neural networks with uncertain target positions

A curious side-issue of the microcalcification problem arose when we noticed that the coordinates given for the microcalcifications frequently did not match their apparent positions in the mammograms. Although this needs to be addressed by the radiologists who provide the data, it raises the interesting problem of how we should train a network in such circumstances. We developed two possible objective functions for this problem, with the usual argument for the cross-entropy error function as a model. This argument interprets the output of the network as the probability that a target is present, conditioned on the input vector. If this probability indicates that the input vector is a positive example, then minimizing the cross-entropy error gives the network that is maximally likely to produce the desired outputs in the training data, given the input vectors.

In the first approach the network is trained so that it is maximally likely to produce the correct output, i.e., a positive response at each of the target positions and a negative response elsewhere. However we don't know the correct target positions and so must average over them. This gives the error function

$$E = -\sum_{i \in \text{Positives}} \log\left\langle \frac{y(\mathbf{f}_i)}{1 - y(\mathbf{f}_i)} \right\rangle_{\pi_i} - \sum_{\text{All } x} \log(1 - y(\mathbf{f}(x))) \qquad \text{(EQ 2)}$$

[1] The research on mammography was largely funded by The Murray Foundation. Follow-up funding under the auspices of the National Information Display Laboratory (NIDL), for which Sarnoff is the host institution, has been approved, but has not yet started.

Invited talk at ONR Sensor Fusion Workshop at Woods Hole

Invited talk entitled "Combining Neural Models and Feature Pyramids for Sensory Fusion".

Congressional Exposition "New Frontiers in Breast Cancer Research"

Presented neural network/pyramid architecture at Congressional exposition entitled "New Frontiers in Breast Cancer Research". The material presented illustrated the dual-use application of our NN/PYR architecture (ATR and mammography). Our work received wide media coverage with write-ups in the Wall Street Journal, and coverage on "CBS This Morning."

Summary of Substantive Information Derived from Special Events:

At the ARPA Image Understanding workshop, we spoke with Thomas Purcell of Booz-Allen Hamilton, who works with NPIC. They are in the definition stage for a program called BEACON which is to identify and transfer technology which will support their image analysts' needs.

Problems Encountered and/or Anticipated:

None

Action Required by the Government:

The most recently scheduled funding increment has not occurred.

Financial Status

1. Amount currently provided on contract: $225,740
2. Expenditures and commitments to date: $251,624
3. Funds required to complete work: $451,130

Accesion For
NTIS CRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification ....
By ....
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

in which $\mathbf{f}(x)$ is the input feature vector at position x, $y(\mathbf{f})$ is the network output given f, π_i is the probability over positions of finding the i-th positive example, and $\langle \ldots \rangle_{\pi_i}$ indicates an average over positions, weighted by π_i. The derivation of this equation assumes that the distributions π_i do not overlap each other. A more general equation which allows for overlapping distributions can be derived, but its use is less convenient.[2]

In the second approach, we want to train the net so that it is maximally likely to produce at least one positive response within each positive region, and a negative response at all locations outside of the positive regions. Two key differences between the two approaches are (1) we do not use probabilities over positions and (2) though overlaps are possible, they do not affect the error function's form. The resulting error function is

$$E = -\sum_{x \in \text{Negatives}} \log(1 - y(\mathbf{f}(x))) - \sum_{i \in \text{Positives}} \log\left[1 - \prod_{x \in i} (1 - y(\mathbf{f}(x)))\right] \qquad \text{(EQ 3)}$$

We trained networks using EQs 1 and 3 on the microcalcification problem. ROC curves indicate that network accuracy is similar for the different error functions. However, those trained using EQ 3 had outputs which are more consistent with the conditional probability interpretation. Specifically, the performance is not very good at low resolution, and one would expect the detection probability to be near zero, since the network should never be very certain that a microcalcification is present, and the a priori probability is very low. The network trained using EQ 3 produced low ouputs, whereas the network trained using EQ 1 had an output near 1 for many examples, including many false positives.

<u>IU Workshop Paper Presented:</u>

We presented the Image Understanding workshop paper entitled, "Neural Network/Pyramid Architectures That Learn Target Context", at the November 1994 Image Understanding Workshop (see attachment to the last quarterly report).

<u>NIPS Poster Presented</u>

Poster presentation entitled "Coarse-to-Fine Image Search Using Neural Networks" at the Neural Information Processing Systems Conference in Denver, CO, on November 30. We will write a paper for the proceedings.

<u>Talk Presented at NIPS Workshop on Neural Networks in Medicine</u>

Invited talk entitled "A Dual-use Neural Network/Pyramid Architecture for Learning Image Context in Mammography".

[2]Unfortunately, overlapping distributions are very common, especially in low-resolution images.